Crystal structure of *Escherichia coli* HdeA

While attempting to solve the crystal structure of cyanovirin-N[1], we inadvertently obtained the structure of an unrelated *Escherichia coli* protein, HdeA (where Hde stands for hns-dependent expression), which we subsequently found to be the major component of the particular sample used for crystallization. Our experience may be an indication of what is to be expected when 'structural genomics' projects aimed at solving protein structures without regard to their function, but rather in search of new structural motifs[2] will be done by design rather than by chance.

Crystallographic procedures are summarized in the legend to Fig. 1. The Patterson self-rotation function calculated with native diffraction data showed clearly the presence of noncrystallographic two-fold symmetry. The structure was solved by multiple isomorphous replacement (MIR), using four heavy atom derivatives (Table 1). Based on the MIR-phased electron density maps, a partial polyalanine structure model of about 70% completeness was built for the two, mostly helical, protein monomers related by noncrystallographic symmetry. However, the electron density maps showed an additional block of density that did not belong to either of the monomers, and the traced structure of the monomers formed noninteracting layers. This result clearly indicated the presence of a third monomer, located near the crystallographic two-fold axis and forming a dimer similar to the noncrystallographic dimer formed by the first two monomers.

At this stage of model building, we realized that the map could not represent cyanovirin-N, especially after we analyzed the NMR-derived coordinates, which contained no helical secondary structure[3]. Because the identity of the protein that we had crystallized was not immediately known, we built a trial sequence based on the shapes and locations of the side chain densities. The resulting structure was subsequently refined (an R_{free} of 32.0% and an R-factor of 25.5%). The presence of a cluster of three aromatic side chains, one of them clearly a tryptophan, adjacent to a disulfide bridge, together with mass spectrometry data on the crystallization sample, which showed that the main component had an M_r of 9,739.7, provided an initial indication that the structure was that of HdeA. This protein is only slightly smaller than cyanovirin-N. Further crystallographic refinement, utilizing the sequence of HdeA, resulted in a model containing 76 residues (10–85) out of 89. The N- and C-terminal residues were significantly disordered in all three molecules, and their positions could not be determined from the electron density maps. The statistics of the final refined structure are summarized in Table 1.

The structure of HdeA is a single domain consisting primarily of an up-down-up-down helical bundle (Fig. 1*a*). Helix A (residues 18–22) is short, whereas the other three helices, B (29–39), C (52–67), and D (74–83), are longer. The four helices were separated to two segments by the tip of the extended loop connecting helices B and C (41–44). The two segments are connected by a disulfide bond (Cys 18–Cys 66) located between helices A and C. Both termini of each molecule are disordered.

Both the crystallographic and noncrystallographic dimers of HdeA are essentially the same. The dimer is formed by extensive interactions among helix B, linker BC, and the observed part of the N terminus of each molecule (Fig. 1*b*). The helices B and B' are nearly parallel and unusually close to each other. The interactions within the dimer interface are mostly hydrophobic and the surface area that was buried upon dimerization is ~850 Å^2. Due to the unusually strong and close dimer interactions, dimeric structure may be essential for the function of HdeA.

Extensive similarity searches based on the nucleotide and protein sequences of HdeA failed to reveal any significant matches with sequences from other organisms. Indeed, it appears that the

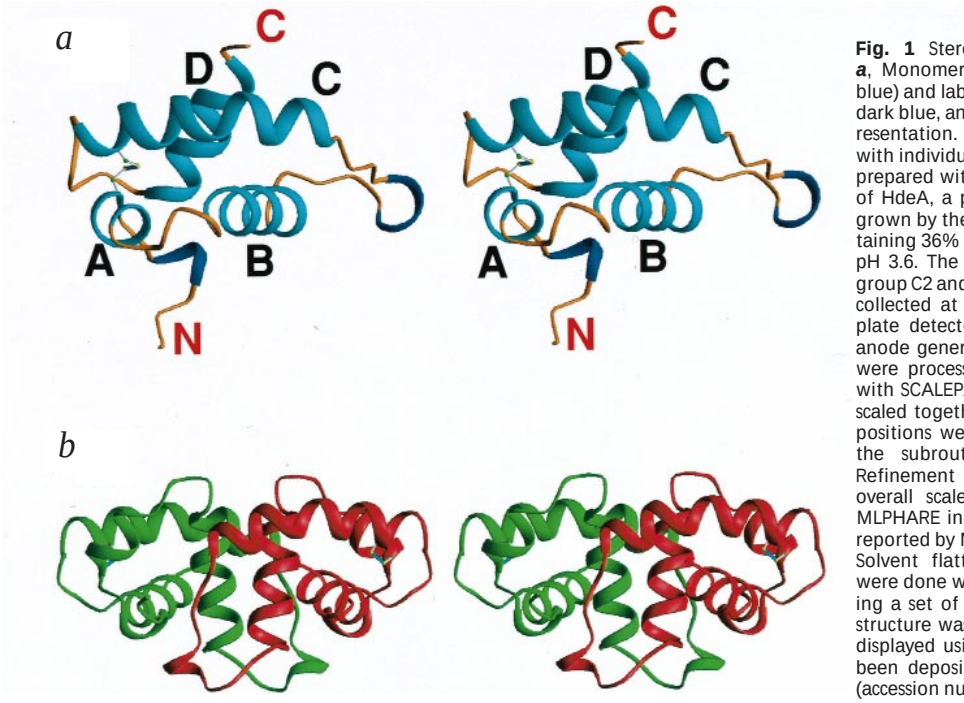

Fig. 1 Stereo diagrams of the structure of HdeA. ***a***, Monomer of HdeA, with marked helices A–D (light blue) and labeled termini. The helical turns are marked in dark blue, and the disulfide is shown in ball-and-stick representation. ***b***, A noncrystallographic dimer of HdeA, with individual molecules shown in red and green. Figure prepared with the program RIBBONS[9]. Methods: Crystals of HdeA, a protein consisting of 89 amino acids[6], were grown by the hanging drop method from a solution containing 36% PEG400, 5% glycerol, 50 mM sodium citrate, pH 3.6. The rod-shaped crystals belonged to the space group C2 and diffracted to 2.2 Å. All diffraction data were collected at room temperature with a MAR345 image plate detector mounted on a Rigaku RU-200 rotating anode generator, operated at 50 kV and 100 mA. Data were processed using the program DENZO and scaled with SCALEPACK[10]. Native and derivative data sets were scaled together using the CCP4 programs[11]. Heavy atom positions were determined by Patterson searches using the subroutine HASSP[12] of the program SOLVE[13]. Refinement of the heavy atom parameters and of the overall scale and temperature factors was done by MLPHARE in the CCP4 suite. The overall figure of merit reported by MLPHARE before solvent flattening was 0.55. Solvent flattening and noncrystallographic averaging were done with the program DM in the CCP4 suite, yielding a set of phases with a figure of merit of 0.76. The structure was refined with X-PLOR[14] and the maps were displayed using the program O[15]. The coordinates have been deposited in the Brookhaven Protein Data Bank (accession number 1bg8).

entire *hde* operon is quite specific to *E. coli*. A comparison with other three-dimensional structures of proteins by using DALI[4] showed a limited similarity between HdeA and almost the entire small N-terminal domain of the *Salmonella typhimurium* chemotaxis receptor methyltransferase CheR[5]. Superposition of 50 Cα coordinates of the two proteins, corresponding primarily to three helices (B–D) in HdeA, yielded a root-mean-square (r.m.s.) deviation of 2.6 Å. Although the topology of these helices is similar in the two structures in terms of their sequential location, we cannot rule out that the similarity is coincidental, especially since it is not based on any sequence similarity (only 3 amino acids out of 50 are identical, a result expected by chance). In addition, the quaternary structures of the two proteins are quite different: the small domain of CheR makes extensive contacts with the large domain, whereas HdeA is a homodimer; the surface of the small domain of CheR is predominantly positive, while that of HdeA is negative; and a cofactor-binding site is formed almost exclusively by the large domain of CheR, with only a minor contribution from the small domain, thus providing no guidance to the interpretation of possible binding of other molecules by HdeA. Thus, any apparent similarity of the fold of these two proteins gives few, if any, clues to the biological significance of HdeA.

Little is known directly about the biological role and the importance of HdeA in *E. coli*. It could be a periplasmic or extracellular protein as indicated by the presence of a signal peptide sequence in the HdeA precusor. The gene encoding HdeA was initially identified as part of an operon regulated by the nucleoid protein H-NS[6]. Recent studies of *E. coli* mutants in which the *hdeA* gene was either deleted or modified by insertion gave different results, showing the hdeA gene to be either essential or nonessential, depending on the nature of the insertions[7]. The results with the deletion mutations, however, indicated that the gene was not essential under the conditions of the experiments .

Another case of mistaken identity of a crystallized protein has been reported[8]. The structure of lobster muscle enolase was solved unintentionally — the structure of lobster arginine kinase was the target. However, this was a significantly different case, because enolase had known structure and function. Based on their unique amino acid sequences and location in an uncharacterized locus on the *E. coli* genome, proteins encoded by the *hde* operon may indeed fulfill some of the requirements for targets of structural genomics projects. The difficulty in assigning a function for HdeA after its structure has been solved is a warning that interpreting the results of future projects in structural genomics may not be straightforward.

Table 1 Unit cell parameters, heavy atom data and statistics of the refined structure

Unit cell parameters and heavy atom data

	native	$SmCl_3$	$KAuCl_4$	$SmCl_3$ + $KAuCl_4$	$Pb(CH_3)_3Ac$
a (Å)	47.48	47.76	47.56	48.17	47.73
b (Å)	76.15	75.78	76.47	76.42	76.72
c (Å)	73.61	73.89	73.28	74.03	73.62
β (°)	98.93	99.36	98.84	99.26	98.64
Resolution (Å)	2.2	2.35	2.5	3.5	3.25
R_{merge}[1]	7.1%	6.6%	7.4%	8.0%	7.3%
Completeness	97.2%	97.4%	96.3%	56.0%	91.1%
R_{iso}[2]		19.8%	15.2%	27.5%	20.2%
R_{cullis}[3]		0.71	0.82	0.71	0.84
Phasing power[4]		1.61	1.09	1.63	1.15

Statistics of the refined model of HdeA

No. of reflections	11,106
No. of protein atoms in a monomer	587
No. of solvent atoms	182
R-factor[5]	16.3% (10,010 reflections)
R_{free}[6]	27.8% (1,096 reflections)
Resolution	10–2.2 Å
R.m.s. deviation of bonds from ideality	0.014 Å
R.m.s. deviation of angles from ideality	1.78°

[1]Rmerge= $\Sigma_h\Sigma_i|I(h,i)-<I(h)>|/\Sigma_h\Sigma_iI(h,i)$ where I(h,i) is the intensity of the i^{th} measurement of h and <I(h)> is the corresponding value of I(h) for all i measurements.
[2]$R_{iso}= \Sigma|F_{PH}-F_P|\Sigma F_P$
[3]$R_{cullis}=\Sigma||F_{PH}\pm F_P|-F_H|/\Sigma|F_{PH}-F_P|$
[4]Phasing power is the mean value of the heavy atom structure amplitudes divided by the residual lack of closure, as defined MLPHARE[11]
[5]R-factor= $\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$
[6]R_{free} was calculated using subset of 10% randomly selected reflections excluded from refinement

Acknowledgments
We thank M. Clore and A. Gronenborn for the coordinates of cyanovirin-N, L.K. Pannell for mass spectrometry analysis, and A. Arthur for editorial assistance. Research sponsored in part by the National Cancer Institute, DHHS, under contract with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Fan Yang[1], Kirk R. Gustafson[2], Michael R. Boyd[2] and Alexander Wlodawer[1]

[1]Macromolecular Structure Laboratory, ABL-Basic Research Program, and [2]Laboratory of Drug Discovery Research and Development, DTP, NCI-FCRDC, Frederick, Maryland 21702-1201, USA

Correspondence should be addressed to A.W. email: wlodawer@ncifcrf.gov

1. Gustafson, K.R. *et al. Biochem. Biophys. Res. Commun.* **238,** 223–228 (1997).
2. Shapiro, L. & Lima, C.D. *Structure* **6,** 265–267 (1998).
3. Bewley, C.A. *et al. Nature Struct. Biol.* **5,** 571–578 (1998).
4. Holm, L. & Sander, C. *J. Mol. Biol.* **233,** 123–138 (1993).
5. Djordjevic, S. & Stock, A.M. *Structure.* **5,** 545–558 (1997).
6. Yoshida, T., Ueguchi, C. & Mizuno, T. *J. Bacteriol.* **175,** 7747–7748 (1993).
7. Link, A.J., Phillips, D. & Church, G.M. *J. Bacteriol.* **179,** 6228–6237 (1997).
8. Duquerroy, S., Le Bras, G. & Janin, J. *Proteins Struct. Funct. Genet.* **18,** 390–393 (1994).
9. Carson, M. *J. Appl. Crystallogr.* **24,** 958–961 (1991).
10. Otwinowski, Z. & Minor, W. *Meth. Enz.*. **276,** 307–326 (1997).
11. CCP4:SERC Collaborative Computing Project No. 4. (Warrington, UK; 1979).
12. Terwilliger, T., Kim, S.-H. & Eisenberg, D. *Acta Crystallogr.* **A43,** 1–5 (1987).
13. Terwilliger, T.C. & Berendzen, J. *Acta Crystallogr.* **D52,** 749–757 (1996).
14. Brunger, A. *in X-PLOR Version 3.1: A system for X-Ray crystallography and NMR* (Yale University Press, New Haven, Connecticut; 1992).
15. Jones, T.A. & Kieldgaard, M. *Meth. Enz.* **277,** 173–208 (1997).